UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

UNITED STATES OF AMERICA

-v-

JOSHUA ADAM SCHULTE,

Defendant.

S3 17 Cr. 548 (JMF)

DECLARATION OF STEVEN M. BELLOVIN, Ph.D.

STEVEN M. BELLOVIN, Ph.D., declares under penalty of perjury:

Personal Background and Qualifications

1. I am the Percy K. and Vida L.W. Hudson Professor of Computer Science and affiliate law faculty at Columbia University. I was retained by defense counsel in this case to assist as an expert in computer systems and computer security. I was present at the criminal trial against Mr. Schulte and heard the testimony of the government's experts, Mr. Leedom and Mr. Berger. I make this declaration in support of Mr. Schulte's motion for access to material not made available to him before his first trial. This declaration is based on my personal knowledge and more than 50 years of experience as a computer expert.

2. I have held a Top Secret/SCI clearance for about 15 years, for my service on a Department of Homeland Security advisory committee, some of my work with the National Academies of Science, Engineering, and Medicine, and my employment during 2016 at the Privacy and Civil Liberties Oversight Board (PCLOB). To perform my work at PCLOB, I needed to review highly classified intelligence programs.

3. My curriculum vitae is annexed hereto as Exhibit "A." In summary, I received my

doctorate in computer science in 1982 from the University of North Carolina at Chapel Hill. I am currently the Percy K. and Vida L.W. Hudson Professor of Computer Science at Columbia University and an affiliate faculty member at Columbia Law School. I have been a Professor of Computer Science at Columbia since 2005. I have also worked as Chief Technologist for the Federal Trade Commission (2012-2013), Adjunct Professor of Computer Science at the University of Pennsylvania (2002-2004), member of the technical staff at Bell Labs and AT&T Labs Research (1982-2004; I was named an AT&T Fellow in 1998), and as a part-time consultant for AT&T (2005-2012).

4. I am also a member of the National Academy of Engineering ("National Academy") and have served on many National Academy study committees and the National Academy's Computer Science and Telecommunications Board. I have also been part of the leadership of the Internet Engineering Task Force, serving on the Internet Architecture Board and as a Security Area Director. I have also served on several advisory committees at the Department of Homeland Security and the Election Assistance Commission.

5. I have published extensively on a wide range of subjects relating to Internet security, computer science, forensics, and network intrusion detection.

Definitions

6. There are several types of data at issue here:

   a. Mirror Images: A mirror image is a bit-for-bit copy of a disk. A mirror image copy includes all: (i) files, (ii) associated metadata (include "access control lists" that say who has what access rights to a file), and (iii) "free space," meaning areas of the disk not currently used by any file, but which may have been used by now-deleted files. A mirror image is used by an expert to run analytic commands on files on a disk.

b. Forensic Image or Forensic Case: A forensic image is standard forensic software, such as the AccessData Forensic Toolkit used by the government. A forensic image extracts and indexes files from a mirror image or from "free space" on the disk.

c. Backup files: A backup file is similar to a .zip file. It is a single file containing multiple files from a disk, which are typically stored elsewhere.

Facts and data do not equal mirror images; if they did, that is what the CIA would have given its expert.

7. Before the first trial, the government granted its forensic expert, Patrick Leedom, access to the mirror images of Mr. Schulte's CIA workstation, the CIA's ESXi server, and the NetApp server (also known as the Altabackup or FS01 server). I was never provided with the mirror images of Mr. Schulte's workstation or the relevant servers. Instead, I received select files from the three servers (including from virtual machines hosted on the ESXi server) and redacted copies of the March 3, 2016 and March 4, 2016 Confluence backup files.

8. The government claims that Mr. Schulte has not been prejudiced by its decision to grant its forensic expert Patrick Leedom—but not me or defense counsel—access to the "mirror images" of the Schulte Workstation, ESXi Server, and NetApp Server. The government asserts that the defense has had all of the information upon which Mr. Leedom relied to arrive at his opinions well before trial, and thus a reasonable opportunity to test and scrutinize those opinions.

9. The government's assertions are not only incorrect, they are also based on the faulty premise that facts and data compiled by an adversary's expert are sufficient, and that a defense witness may be forced to base his expert review and opinion on information culled by the adversary's expert. As explained below, access to the original data is paramount, which is why the CIA provided access to Mr. Leedom.

10. First, while the government has repeatedly stated that it has made a substantial amount of forensic material available to the defense, it is substantial only in the abstract. Compared to the data made available to its own expert, the production to the defense is miniscule and far from complete. In no other case has my review been so limited in scope.

11. Second, because I have never been granted access to the mirror images of the ESXi and FS01 Servers, I have been unable to reproduce (and thereby confirm or refute) all the analyses and tests that Mr. Leedom was able to perform. This hindered my ability to testify and Mr. Schulte's ability to properly cross-examine Mr. Leedom.

12. Third, without access to the full mirror images, I have been unable to conduct my own forensic examination of the data, based upon my own training and experience, and to explore various hypotheses, data, and conduct additional analyses that Mr. Leedom may have missed or misinterpreted.

13. Without a doubt, the material provided cannot be utilized to conduct an adversarial forensic examination to put Mr. Leedom's analyses to the test through the scientific method.

Tests and Analyses of Stash and Confluence backups

14. As an initial matter, there are two classes of data that the government claims were stolen, a

backup of the Confluence application (select materials purportedly published as "Vault 7" by WikiLeaks) and a backup of the Stash application (select materials purportedly published as "Vault 8" by WikiLeaks), which were stored on the FS01 server along with multiple other backups. Confluence is in essence a "wiki", a collaboratively edited set of documents created and used by DevLAN personnel. Stash is a front-end for "Git", an open source "version control system." A version control system is similar to the "track changes" facility in Microsoft Word, but far more powerful. It can permit retrieval of files as they existed at any point in the past; it also permits one group of developers to work on, say, Version 2.0 of a system while others work on 1.1 and still others work on Version 1.0.1 to fix bugs in Version 1.0. All of these versions are stored simultaneously by Git (and hence by Stash).

15. The government provided the defense redacted Confluence backups from March 3, 2016 and March 4, 2016. The government removed all CIA content from these backups, so the defense was unable to perform any analyses or tests.

16. Several months before the first trial, I was given just several hours at a CIA facility to examine a *redacted* version of a few backup files, with no access to Internet reference material. No working programmer functions without full access to online documentation. Indeed, during my own time working in a SCIF at PCLOB, I had two separate computers, one for classified work and one for general internet work. Such dual access was provided to CIA developers and almost certainly to Mr. Leedom—if nothing else, he could go back to his office to look things up—but not to me.

17. One of the major analyses performed by Mr. Berger that I could not reproduce without access to the stash and confluence backups stored on the FS01 Server was his "timing analysis" (Tr. 1351-52). Mr. Berger relied upon all backups to note each day particular files

were modified, and which of these versions were ultimately released by WikiLeaks.

[redacted]

18. Additionally, Mr. Berger could not have performed his timing analysis for the Stash or Confluence backups without a mirror image of the FS01 server, or at the very least, access to all the Stash and Confluence backup files in order to access and compare different versions of the backups to the leaked files.

19. [redacted]

20. Mr. Leedom asserted at trial that the March 3, 2016 Confluence backup contained a SQL file with a "character coding error" that also appeared in the Vault 7 and 8 disclosures (Tr. 1118:4-25, 119:1-4). The existence of the error was crucial to Mr. Leedom's conclusion that the leaked data came from a backup as opposed to the Confluence Virtual Machine (Tr. 928:13-17, 1129:13-16). But the error, itself, was not obvious. It took the government analysts weeks to discover it.

21. [redacted]

[redacted]

22. Before the first trial, Mr. Schulte requested information on the access times (the last time a file was used) of certain Confluence backup files. If I had access to a mirror image of the backup server, I could have checked these times myself instead of alerting the government to defense strategy.

23. Access times can be manipulated by a "touch" command. It was clear from the government's subsequent disclosure that Mr. Leedom had not examined the access times before the defense alerted him to the issue, and hence, did not properly investigate whether a touch command had been used by someone to alter them, let alone by whom. The access time issue was an element of defense strategy, one that would have remained unknown to the government if I had had the same access to the backup server as Mr. Leedom (GX 1207-27 and 1207-30). The defense was thus doubly hindered: it could not properly investigate this point, and its strategy was disclosed prematurely.

24. At the first trial, the government sought to prove the exact date and time when Mr. Schulte allegedly accessed and stole the March 3, 2016 Confluence and Stash backups: April 20, 2016 at 5:35-6:51 pm (Tr. 1063:3-12). Because I was not able to examine the backup server, I could not verify or rebut the government's assertion. On cross-examination of Mr. Leedom (and in summation), Ms. Shroff raised the possibility that a "touch" command was used to change the access time, and Mr. Leedom acknowledge that such a command can be employed to alter time stamps (Tr. 1181:16-25, 1182:1-6). [redacted]

[redacted]

[redacted]

[redacted]

[redacted]

28. [redacted]

29. [redacted]

30. [redacted]

31. The government also asserts that it has provided large amounts of information to the defense, including log files, but this data is in no way equivalent to the information to which Mr. Leedom has had access.

32. The government claims that it gave the defense log files from the ESXi server. In fact—per the defense *ex parte* letter to the Court of February 12, 2019—some of those files were demonstrably damaged. While the government now asserts it has corrected this issue, the

underlying question of the integrity of the mirror images remains; without the mirror images, the defense cannot confirm whether the government followed proper forensic protocols or otherwise tainted and destroyed relevant data.

33. The government claims that it provided the Confluence databases to the defense. But those databases appear to have been heavily redacted, with all content files either missing or deleted, including those allegedly released by WikiLeaks. Furthermore, they did not include the apparently damaged "SQL" file that Mr. Berger evidently used in his analyses. GX 1704, Slide 3 (GX 1207-97) [redacted]

34. The government claims the defense had the "backup script" in sufficient time to determine whether Mr. Leedom's claims about the damage to the SQL file are accurate. In fact, without access to the mirror images, and for reasons too complex and technical to explain here, the backup script alone does not permit the defense to assess the validity of all of Mr. Leedom's assertions.

35. The government claims that it has produced to the defense all of the unallocated space from the ESXi server "about which Mr. Leedom testified." But "the unallocated space …about which Mr. Leedom *testified*" (emphasis added) is not the same as all the unallocated space he *examined.* [redacted]

36. At trial, Mr. Leedom testified about Mr. Schulte's encrypted private SSH key. But Mr. Leedom did not verify that Mr. Schulte's private key corresponded to the public key file,

and I was not able to do it myself because I lacked access to machine-readable and processable copies of the files purported to be Mr. Schulte's private and public SSH keys.

37. Upon examining the Stash backups, I will also need sufficient Internet access to check on Git operational details and other reference materials necessary for detailed examinations. Git is notoriously complex; indeed, its complexity is regularly the subject of joke in the technical community, e.g. https://xkcd.com/1597/.

38. [redacted]

39. Finally, I note that forensic examiners are not infallible. The only way I can test if the government made a mistake in their analysis, missed critical data, or otherwise misinterpreted data is through equal access to the mirror images and adversarial testing. In all my experience, including civil litigation and a criminal prosecution in which I worked with the FBI, they always provided me with a mirror image copy of the suspect's disks.

I declare under penalty of perjury that the foregoing is true and correct.

April 22, 2022

Steven M. Bellovin, Ph.D.